MARYLAND DEAF CULTURE DIGITAL LIBRARY

The Future Is Multimodal

Ensuring Access for Deaf Communities in a Voice-First World

MD Tech Connect 2025 | Technology As A Community Building Tool

Keith Delk (He/Him)
Program Specialist

Brandt Van Unen (He/Him)
Library Coordinator/Librarian

Key Takeaways for Today

The goal of this session is to give you a framework on how to:

- **Evaluate new technologies** to see whether they offer *true* Functional Equivalence or just basic access
- **Recognize the limitations of voice-first tools** and how designing for multimodal accessibility from the start leads to better outcomes.
- **Build AI literacy through a Deaf-centered lens,** recognizing how language, culture, and communication norms shape equitable AI use.

Context and the Gap

The Voice-First Era

The world is being optimized for audio with AI-driven technologies such as smart speakers, conversational interfaces, and smart wearables

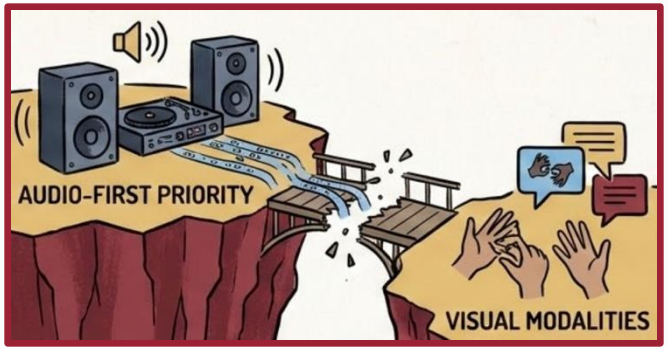
- Voice assistants are often audio-only.
- Real-time translation focuses on spoken languages.
- **"Voice-first = Exclusion-first"** for Deaf communities

Digital Phonocentrism

Phonocentrism is the belief that speech is inherently superior to visual modalities like written and sign language

Digital systems listen, but don’t see

Technological design creates a digital divide by prioritizing audio input



The Gold Standard

It’s not about sameness; it’s about equity of experience, or **Functional Equivalence.**

Core question for technology:

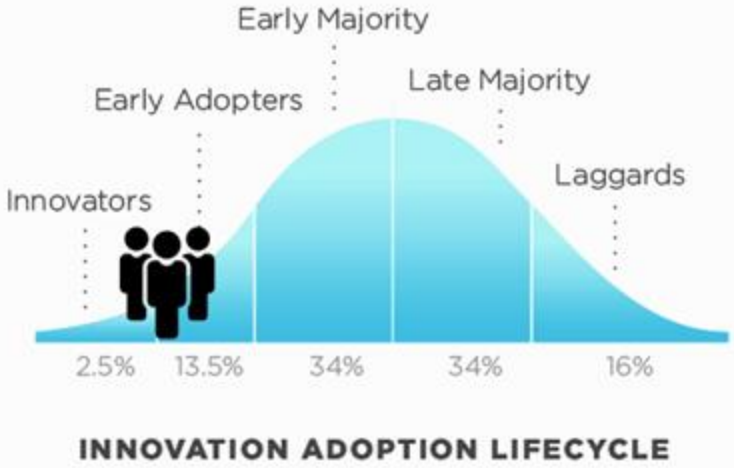
Can a Deaf person achieve the same goal with the same level of effort as a hearing person?

The Technological Tug of War

Deaf Communities are constantly forced to wait for accessibility to be retrofitted onto existing technology.

- Cable TV -> Decoders added later
- Telephony -> TTY/VRS added later
- YouTube -> Craptions added later
- Voice Assistants -> ??

The Innovation Chasm



Where We Want To Be

- Early Adopters, trying new tools tha facilitate communication

Where We Are Forced To Be

- Laggards, waiting for accommodations to be retrofitted

Goal: Design inclusively and multimodally from day one.

AI's Promise and Shortcomings

Speech to Text (STT)

The Promise

- If you can speak, the machine understands
- Efficiency

Reality

- Most ASR models are trained on “standard”, “normative” speech patterns

Text to Speech (TTS)

The Promise

- Audible voice for non-speaking and sign language users
- Audio access
- Autonomy & Amplification

Reality

- Audiophiles are generally against AI voices
- Struggles with intonation
- Audio-Centric
- “Deaf Authoring Tax”

AI/Auto Captioning & Transcription

The Promise

- Fast & on demand
- Scalability
- NSI & Contextual/Expressive

Reality

- Privacy laws
- Inaccurate, leading to non-compliance
- Text-first

AI-assisted human captioning is promising

The Pivot Point

AI-Assisted Accessibility for the Deaf

Sign Language Detection (SLD)

Detects when signing is happening

Sign Language Recognition (SLR)

Converts sign to text in real-time

Sign Language Translation (SLT)

Translates between written/audio content and sign

Case Study: The Promise of Text-to-Video - Voice

Prompt: A person using a hologram to make a video call while walking in a digital utopia, saying “Hey, how's it going? Just stepping into the future.”

Case Study: The Promise of Text-to-Video - ASL

Prompt: A person using American Sign Language with a hologram for a video call in a digital utopia, saying “Hey, how's it going? Just stepping into the future.”

Let's try again… but with parameters of ASL

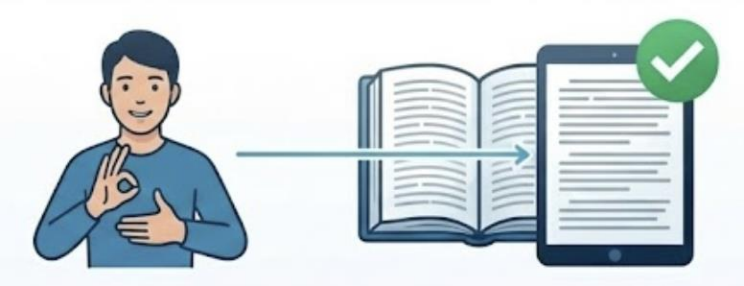
"A person signing 'Deaf' touching the tip of dominant forefinger near the ear and then moving the forefinger to the tip of mouth."

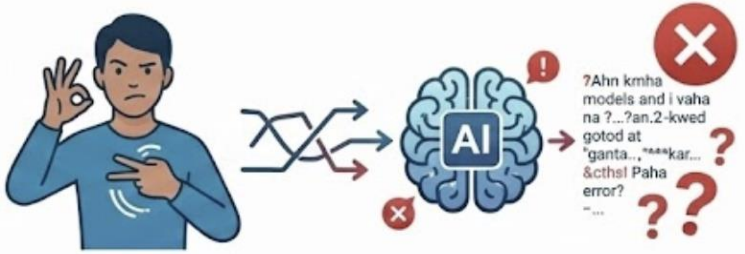
A woman trying to sign the word, “Deaf”

TL;DR

- Text-to-Video generate stunning visuals but struggle with precise linguistics of Sign Language.
- They create inaccurate lip-syncing, expressions, and movements that look like signs but lack grammatical meaning.

Sign Language/Video to Text (V2T)





The Promise

- Automatic captions from ASL videos
- Replace voice interfaces
- Make sign language searchable

Reality

- Trained on lab-created datasets, not natural sign language
- Recognize isolated signs, not fluent connected signing

Text to Sign Language/Video (T2V) or Sign Avatars

The Promise

- Faster content creation
- On-demand ASL Translations
- Accessibility at scale

Reality

- Trained on very limited SL datasets
- Not grammatically accurate

AI Tools for the Deaf - TL;DR

- Text-to-Video generates visuals but struggles with sign language linguistics.
- AI for sign language often creates output that is grammatically inaccurate.
- Video to Text models are trained on isolated signs, not natural, fluent communication.

AI Tools are great for access to information.

The "Holy Grail" of Deaf accessibility is and always will be human-based interpretation for human communication

Next time you evaluate software for your library, ask yourself: 'Does this interface support multimodal access to the same information?”

MARYLAND DEAF CULTURE DIGITAL LIBRARY

AI Literacy: Deaf Culture Edition

The future is about to be here. Are we ready?

Current and Possible Issues

AI-Generated Scam Calls

VRS/VRI Interpreter's Role

- Relay messages

Current Defense

- Number Blocking

Limited Data and Biases

Critical challenges in AI development for Deaf communities.

Who Feeds The Information

- Lack of diverse, authentic input sources.

Insufficiency (Too Few Signers)

- Datasets underrepresent signer population.

Inaccuracy (Regional Signs, Gestures, Etc)

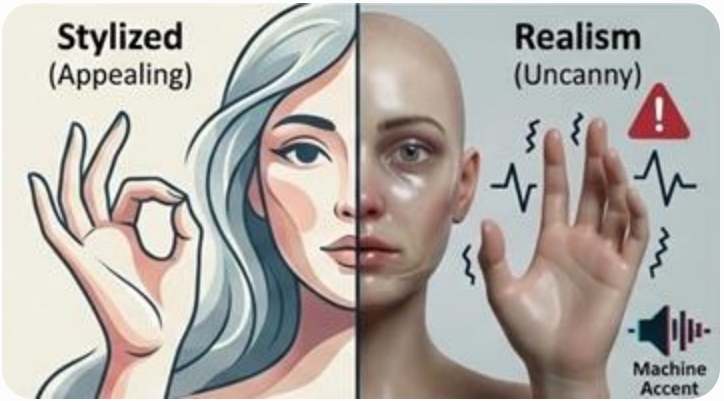
- Variations in language lead to errors.

Ai Pushes Back On "Correct Signs"

- Algorithms favor standardized over real-world signing.

Signing Avatars

- Replacing Librarians, Teachers, Instructors, etc
- Uncanny Valley Effect
- “Realism” vs. “Motion”
- Facial animations and lip sync
- Realism vs. Stylized
- “Machine Accent”



Reasonable Accommodations

- **ADA is in need of being updated**
  - Technological advancements
  - Gaps in protection
- **Interpretations of what is "reasonable”**
  - Minimum standards
  - Guidelines
  - Best as possible
  - Excessive demands

AI-Generated Captioning/Translating

- **National Deaf Center's survey**
  - 87% of deaf college students 'agree' or 'strongly agree' that human-generated captions are more effective than AI-generated ones.
- **AI struggles to 'read' the nuances**
  - Nuances in context, sarcasm, idioms are often missed.
- **Background sounds being misinterpreted**
  - Laughter, music, and non-speech sounds cause errors.
- **Censorship: Double Standards**
  - Inconsistent censorship, blocking important context.

De(?)-Evolution of Sign Language

- VRS/VRI-affected Sign Language
- Al-affected Sign Language
- We accommodate Al or Al accommodate us
- Is creativity worth the price of using Al?

Best Approach with the Issues

Hybrid Model (Human and AI)

When in doubt, contact the DCDL

Resource 1

Resource 2

Resource 3

In Conclusion…

The Future is Multimodal When We Design Inclusively

The “voice-first” world is a temporary and incomplete vision. The future is multimodal. It’s voice, **AND** text, **AND** sign, **AND** tactile.

Emphasize user-centric design. To ensure equitable experiences, you must design and build **with the community.**

Invest in AI literacy for everyone to navigate this future safely and effectively.

The Multimodal Future We Can Build Together

QUESTIONS?

Contact us

Maryland Deaf Culture Digital Library

www.marylanddcdl.org

info@marylanddcdl.org